Effects of dialect and talker variability on lexical recognition memory

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation *with distinction* in Linguistics in the undergraduate colleges of The Ohio State University

by

Terrin N. Tamati

The Ohio State University
June 2008

Project Adviser: Professor Cynthia Clopper, Department of Linguistics

Abstract

The current study investigated recognition memory for dialect variation in a recognition memory experiment with separate training and test phases. In the training phase, participants were asked to identify words spoken by three female talkers from the Midland dialect region and three female talkers from the Northern dialect region. In the test phase, participants listened to another set of words and were asked to indicate whether each word was from the training phase, “old,” or completely new, “new.” In this phase of the experiment, half of the words were “old,” having been previously introduced in the training phase, and half were “new,” not having been introduced in the training phase. Of the “old” words, one-third were repeated by the same talker, one-third were repeated by a different talker from the same dialect region, and one-third were repeated by a different talker from a different dialect region. Based on previous research, it was expected that, for each original dialect, participants would be the most accurate and quickest for the “old” words that were repeated by the same talker, the least accurate and slowest for the “old” words that were repeated by a different talker from a different dialect region, and somewhat in between for the “old” words repeated by a different talker from the same dialect region. The results of this study indicate that episodic memory traces of spoken words retain fine-grained surface details, as found in Goldinger (1996) and Palmeri *et al.* (1993), as responses to same-talker repetitions were generally more accurate and faster than responses to different-talker same-dialect and different-talker different-dialect repetitions. In addition, response time patterns suggest that both abstract lexical representations and episodic traces are stored in long-term memory and contribute to perception. Finally, the significant vowel interactions provide some evidence that dialect information is implicitly coded by the listener, though further studies are needed to better understand this result.

Introduction

The speech signal is robust to individual differences across speakers. Individual differences in attributes, such as vocal tract length (Fant, 1973), dialect (Peterson & Barney,1952; Hillenbrand et al., 1995), and methods of articulation (Ladefoged, 1980), are reflected in the acoustic signal. Nevertheless, without much difficulty, we are able to understand each other. To account for this variation, many theories of speech perception assume that idiosyncratic features of words are "normalized" in relation to canonical mental representations; variation is stripped away to leave only information that is crucial to the identification of the word (e.g., Green, Kuhl, Meltzoff, & Stevens, 1991; Summerfield & Haggard, 1973; Pisoni, 1997). Such "normalization" processes would preclude surface voice information from being stored in long-term memory. However, listeners are able to use the material which would be considered tangential to word identification in communication (Ladefoged & Broadbent, 1957). Indeed, we can get information about a speaker's identity (Van Lancker, Kreimer, & Wikens, 1985; Van Lancker, Kreimer, & Wickens, 1985) and the speaker's origin and background (Labov, 1972) from these idiosyncratic voice attributes. The use of this information indicates that we are actually storing detailed voice information.

In a series of experiments, Goldinger (1996) and Palmeri et al. (1993) investigated recognition memory for spoken words. Palmeri et al. (1993) looked at recognition memory as a function of the number of intervening words between the first and second presentations of a word, the number of talkers in a stimulus set, and voice (whether the word was repeated in the same or a different voice). When listeners were asked to respond to whether the word was repeated or new, same-talker repetitions were recognized more accurately and quickly than different-voice repetitions. Palmeri et al. (1996) also examined the effect of gender, whether the word was repeated by a talker of the same or different gender, on recognition memory. Listeners were faster and more accurate for same-talker repetitions than different-talker repetitions regardless of the gender. The advantage for same-talker repetitions in these studies suggests that participants were sensitive to fine-grained, talker-specific details in the acoustic signal. Later, Goldinger (1996) looked at explicit and implicit memory for spoken words as a function of talkers' voices, time between study and test sessions, and levels of processing. In two experiments, listeners were better at correctly identifying a spoken word when the word was presented to them in the same voice as was previously presented to them. In addition, listeners were found to be sensitive to fine-grained similarity when a word was repeated in a different voice in the implicit task. The results of these studies suggest that this talker-specific information is encoded and retained in long-term episodic memory representations of spoken words.

While dialect is an important source of variation in speech, little is understood about the effect of dialect variation on speech processing. Recent studies have shown that listeners can use dialect-specific acoustic-phonetic details to identify talker dialect (Clopper & Pisoni, 2004). In the Clopper & Pisoni (2004) study, listeners used dialect-specific properties in the acoustic signal to place talkers into dialect categories. Their findings suggest that naïve listeners have knowledge of dialect-specific phonological variation and can use this knowledge to classify talkers.

In a study on cross-dialect lexical decision, Floccia et al. (1996) reported that native French listeners responded more quickly in a lexical decision task when the talker spoke the native French dialect of the listeners than when the talker spoke a non-native French dialect. Moreover, they found that listeners were generally successful when the talker spoke the standard Parisian accent.

This suggests an advantage for both native and standard dialects over non-native and non-standard dialects respectively. Similarly, in a cross-dialect intelligibility study, Labov and Ash (1997) found evidence of an advantage for native dialect listeners in a transcription task under good listening conditions. Listeners from the Birmingham area were more accurate in their responses than listeners from Philadelphia and Chicago when they were presented with phrase-length utterances from a Birmingham talker. In another study, Clopper and Bradlow (in press) reported that the General American dialect, the "standard" American English dialect, was the most intelligible for listeners of all dialect backgrounds at three signal-to-noise ratios (SNRs). Listeners were generally less accurate at transcribing sentences spoken by talkers from the Northern, Southern, and Mid-Atlantic dialect regions. These studies also suggest an advantage for both local and standard dialects and a disadvantage for unfamiliar or more marked dialects.

In Clopper & Pierrehumbert (2007), the interaction of the Midland and Northern vowel systems, and the confusions that result from such interaction, were examined in a word recognition task in noise. While listeners in that study performed similarly for both dialects in terms of accuracy at the vowel level, they made different types of recognition errors for each dialect. For example, listeners confused the vowel /ɛ/, as in *bet*, with the vowel /æ/, as in *bat* significantly more often than chance for the Northern dialect talkers and confused the vowel /ɔ/, as in *caught*, with the vowel /a/, as in cot significantly more often than chance for the Midland dialect talkers. In total, listeners confused /ɪ/ with /ɛ/ , /ɛ/ with /æ/, /ʌ/ with /ɔ/, /ɔ/ with /a/, /ow/ with /ʊ/, and /a/ with /ʌ/ for Northern dialect talkers and /ɔ/ with /a/, /ow/ with /ʊ/, /ʌ/ with /ʊ/, /ow/ with /u/, /ɔ/ with /ʌ/, and /a/ with /ʌ/ for Midland dialect talkers. In that study, listeners responded differently to each dialect, making systematic dialect-specific misinterpretations. The study suggests that some dialect-specific phonetic differences between the Midland and Northern dialects can lead to predictable phonological confusions in isolated word recognition.

These studies all seem to indicate that variability due to the dialect of the talker can interfere with the perception and processing of linguistic aspects of the speech signal. Listeners attend to dialect information in the acoustic signal and can use their knowledge of dialect variation to complete certain tasks, such as dialect classification, but dialect can also interfere with other tasks, such as word recognition and lexical decision. The goal of the current study was to examine how dialect variation is processed and represented in the mental lexicon. To investigate this issue, a recognition memory experiment was carried out. In this experiment, response time and accuracy were examined for the recognition of same-talker, different-talker same-dialect, and different-talker different-dialect repetitions of spoken words. If detailed talker-specific information is retained in long-term memory, performance for same-talker repetitions should be more accurate and faster than when the word is produced by a different talker. If dialect is encoded in the long-term representations of spoken words, then performance should also be affected by dialect matches, where performance for same-dialect repetitions should be faster and more accurate than performance for different-dialect repetitions. Based on the findings of Palmeri et al. (1996) and Goldinger (1993), two main hypotheses arise: 1) listeners' performance will be affected by dialect matches since they are sensitive to such fine-grained similarity when a word was repeated in a different voice, as was found in Goldinger (1996) or 2) listeners' performance will not be affected by dialect matches since gender matches did not affect recognition performance in Palmeri et al. (1993).

RECOGNITION MEMORY EXPERIMENT

In this study, recognition memory for spoken words as a function of the voices of the repetitions was examined. The two-part experiment consisted of a training phase and a test phase. In the training phase, listeners were asked to identify words spoken by three female talkers from the Midland dialect region and three female talkers from the Northern dialect region. In the test phase, participants listened to another set of words and were asked to indicate whether each word was from the training phase, "old," or not from the training phase, "new." Listeners were required to respond only in terms of word identity; they were instructed to classify repeated words as "old" if they had heard it before regardless of whether it was repeated in the same voice or a different voice. In this phase of the experiment, half of the words were "old," having been previously introduced in the training phase, and half were "new," not having been introduced in the training phase. Of the "old" words, one-third were repeated by the same talker, one-third were repeated by a different talker from the same dialect region, and one-third were repeated by a different talker from a different dialect region.

By controlling for the dialect of the talkers, the separate effects of talker and dialect could be assessed. This way, the effects of dialect matches, same-dialect and different-dialect repetitions, and exact talker matches on performance could be compared. Based on the results from Palmeri et al. (1993) and Goldinger (1996), it was expected that participants would be the most accurate and quickest for same-talker repetitions. In consideration of the effects of dialect variability on speech processing in the studies discussed above, it was expected that participants would be the least accurate and slowest for different-talker different-dialect repetitions and somewhere in between for different-talker same-dialect repetitions.

Method

Listeners

Forty undergraduate students at The Ohio State University were recruited to participate in this experiment. Students received credit for either an introductory linguistics or introductory psycholinguistics course for their participation. Data from eight listeners were excluded: four participants were non-native speakers of English, two participants were bilingual, one participant reported a history of a hearing or speech disorder, and one participant was substantially older than the other participants. The remaining 32 participants were all monolingual native speakers of English with diverse residential histories. Twelve lived exclusively in the Midland dialect region before the age of 18, four lived exclusively in the Northern dialect region before the age of 18, two lived exclusively in the Southern dialect region before the age of 18, and the remaining fifteen lived in more than one dialect region before the age of 18. All of these 32 participants were between the ages of 18 and 27. None reported a history of a speech or hearing disorder.

Talkers

Six female talkers from the Indiana Speech Project Corpus (Clopper et al., 2002) were used for this experiment. The Indiana Speech Project Corpus (Clopper et al., 2002) consists of audio recordings of five females for each of six geographic regions of Indiana. Materials for each talker include words and nonsense words read in isolation, read sentences, a read passage, and a conversation with the experimenter. This corpus was designed for acoustic and perceptual studies on regional language variation.

The six talkers used in the current study were monolingual native speakers of American English and were between the ages of 19 and 21 at the time of the recording. Three of the talkers represented the Northern dialect region and three represented the Midland dialect region. The talkers had lived exclusively in their respective regions at least until the age of 18 and had parents who were also from that region. All talkers were from Indiana. Two of the talkers representing the Northern dialect region were from the Chicago suburbs in Indiana and the other talker representing the Northern dialect region was from Fort Wayne, Indiana. The three talkers from the Midland dialect region were all from Indianapolis, Indiana. The map in Figure 1 displays the area of origin of the talkers.

Figure 1: A map of Indiana with the areas of origin of the six talkers indicated with filled circles. Circles represent Chicago suburbs (Northwest corner), Fort Wayne (Northeast), and Indianapolis (Central).

It was anticipated that talkers from these areas would produce dialect-specific variants of the vowels analyzed in this study. For reference, Figures 2 and 3 display the traditional descriptions of the Midland and Northern vowel systems, respectively (Labov *et al*, 2006).

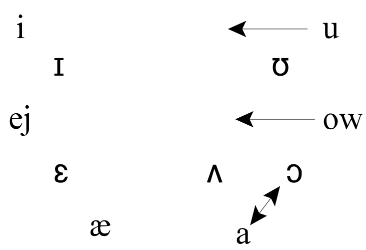

Figure 2: Midland vowel system

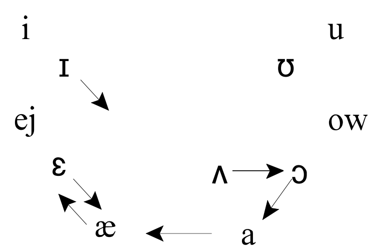

Figure 3: Northern vowel system

The Midland vowel system, as shown in Figure 2, is characterized by a fronted /u/ and /ow/ and the merger of the low back vowels /a/ and /ɔ/. The North vowel system is characterized by the Northern Cities Vowel Shift, which includes the raising and fronting of /æ/, the backing and lowering of /ɛ/ and /ɪ/, the backing of /ʌ/, the fronting of /a/, and the fronting and lowering of /ɔ/.

Stimulus Materials

Stimulus materials were four lists of words produced by the six talkers. All items were monosyllabic words of the type consonant-vowel-consonant (CVC) and came from the Indiana Speech Project Corpus (Clopper et al.,2002). Words were recorded in isolation in a sound-attenuated chamber using a head-mounted Shure microphone (SM10A) and a tube microphone preamplifier (Applied Research Technology). Individual words were digitized at a sampling rate of 44.1kHZ by a Roland UA-30 USB Audio Interface and were recorded into their own .aiff soundfiles on a Macintosh Powerbook G3. For this experiment, words were segmented into individual sound files at a sampling rate of 22050Hz and 16-bit resolution.

Four groups of participants received a different list, each with different training and test words. The lists were constructed from a database of 122 words produced by the three female talkers from the Midland dialect region and the three female talkers from the Northern dialect region. Each group of listeners was presented with a training list that consisted of 66 words. For each of the eleven American English vowels /i, ɪ, ej, ɛ, æ, a, ʌ, ɔ, ow, ʊ, u/, there were six words, which were equally distributed among the talkers. The test list consisted of each of the words in the training list, repeated by either the same talker or a different talker, along with an additional 66 "new" words. The set of repeated words was manipulated so that one-third of all repetitions were produced by the same talker as in the training list, one-third were produced by a different talker from the same dialect region as the original talker, and one-third were produced by a different talker from a different dialect region as the original talker.

The set of "new" words also contained one of the eleven American English vowels /i, ɪ, ej, ɛ, æ, a, ʌ, ɔ, ow, ʊ, u/, but were not controlled for an even presentation of each vowel. Instead, preference was given to "new" words that form a real word minimal pair with the original word from the training phase based upon predicted phonological confusions originating from the traditional descriptions of these two dialects. For example, the word *bat* from a Northern dialect talker might be confused with *bet* since /æ/, in the Northern dialect, is raised and fronted and may be confusable with /ɛ/. Indeed, in this experiment, when *bat* appeared in a training phase list, *bet* was presented as a "new" word. As minimal pairs are only predicted for /ɪ, ɛ, æ, a, ʌ, ɔ, ow, ʊ/ for the Midland and Northern dialects, no minimal pairs could be provided for /i, ej, ow, u/. An average of 19 minimal pairs were presented to listeners over all the lists, each with at least one minimal pair for each of the vowels /ɪ, ɛ, æ, a, ʌ, ɔ, ow, ʊ/. Because of the attempt to include these minimal pairs in the test list, evenly distributing the "new" words across vowels was not possible. Response times and false alarm rates have not yet been examined for these minimal pairs and will not be discussed in this paper.

Each talker's voice, and consequently each talker dialect, was used an equal number of times in each list. Talkers were assigned pseudo-randomly, with each talker represented once for each vowel in the training phase and once for each corresponding "new" and "old" word in the test phase. Minimal pairs were also assigned by these same methods. In the training phase, each talker was presented 11 times and each talker dialect was presented 33 times. In the test phase, each talker was presented 22 times and each talker dialect was presented 66 times. This resulted in each talker being presented a total of 33 times and each talker dialect being presented a total of 99 times in the training and test lists combined.

Procedure

Listeners were seated in individual booths. Each sat in front of a computer equipped with a keyboard, a button box, and headphones. In the training phase, words were presented one at a time over the headphones. Listeners were asked to type the word they heard in a box on the screen. They could take as much time as they wanted to respond, but they could only listen to the word once. Listeners pressed the <ENTER> key to move on to the next trial.

After completing the training phase, listeners were presented with the instructions for the test phase. Once listeners read through these instructions, they could press any button on the button box to continue. Listeners were again presented with words one at a time over the headphones. For the test phase, however, listeners were asked to press a button representing "old" if they had heard the word before, regardless of the voice it was presented in, or hit a button representing "new" if they had not heard the word before. They were asked to respond as quickly as possible without compromising accuracy. Although listeners were previously advised that the experiment consisted of two parts, they were not told the contents of the test phase nor were they instructed to memorize words in the training phase.

The responses from the training phase were manually scored for correct word and correct vowel. For example, a response of *code* for the target word *code* would be scored as correct for both word and vowel, whereas a response of *coat* for the target word *code* would be scored as incorrect for word but correct for vowel. In addition, a one-syllable nonword response containing a vowel judged to be the same as that of the target word would be counted as correct for vowel, as would be the case for *gode*. Multisyllabic words, heteronyms, and nonwords were scored as incorrect. Homonyms (e.g., *knead* and *need*) were scored as correct.

For the testing phase, response times were measured from the onset of each stimulus word. A series of corrections on the response time data were carried out before any statistical analyses were performed. Data replacements were executed according to the methods outlined in Winer (1971). First, data from the extra repetition of three "old" words in the first three lists that was the result of a coding error in the creation of the lists was excluded for all participants in those lists. Second, all response times that were below 500 ms and above 3000 ms were discarded. Third, for each response, two z-scores were calculated, one for subject and one for item. Any responses for which both z-scores were above 3.0 were also discarded. Finally, all discarded values were replaced with the average of the mean value for item and the mean value for subject.

Results

In the following analyses, *Identification Accuracy* refers to the typed responses in the training phase and *Word Recognition Accuracy* refers to the "old" or "new" judgements in the test phase. For the training phase data, *word correct* refers to a response of the same word as the target word and *vowel correct* refers to a response of the same vowel as in the target vowel. In the analysis of response time data, a *hit* was defined as responding "old" to a repeated word. All response time data reported here are for *hits*.

Identification Accuracy

A summary of the identification accuracy scores in the training phase is shown in Table 1.

	Midland	Northern
Word	86 (6)	90 (6)
Vowel	91 (6)	94 (6)

Table 1: Mean percent correct and standard deviation of target words and target vowels for each talker dialect in the training phase.

As shown in Table 1, accuracy at both the word and vowel levels was slightly higher for the Northern dialect talkers than for the Midland dialect talkers. To assess the difference in performance between talker dialects, two separate paired *t*-tests were carried out on accuracy for both word correct and vowel correct with talker dialect as the factor. Prior to analysis, percent correct scores for vowel and word were transformed into Rationalized Arcsin Units (RAU) (Studebaker, 1985). The difference in correct responses for the target word by dialect was significant by a paired *t*-test ($t(31) = -3.06$; $p < 0.01$). However, the difference in correct responses for the target vowel by dialect was not significant by a paired *t*-test ($t(31) = -1.94$; $p = 0.06$).

Word Recognition Accuracy

Mean percentages of hits were calculated for each listener and were used to investigate the effect of the dialect of the talker in the test phase on word recognition accuracy and the effects of repetition type and the dialect of the original talker on word recognition accuracy. Figure 4 displays the mean percent of words correctly identified as "old" by the dialect of the talker in the test phase (Midland or North).

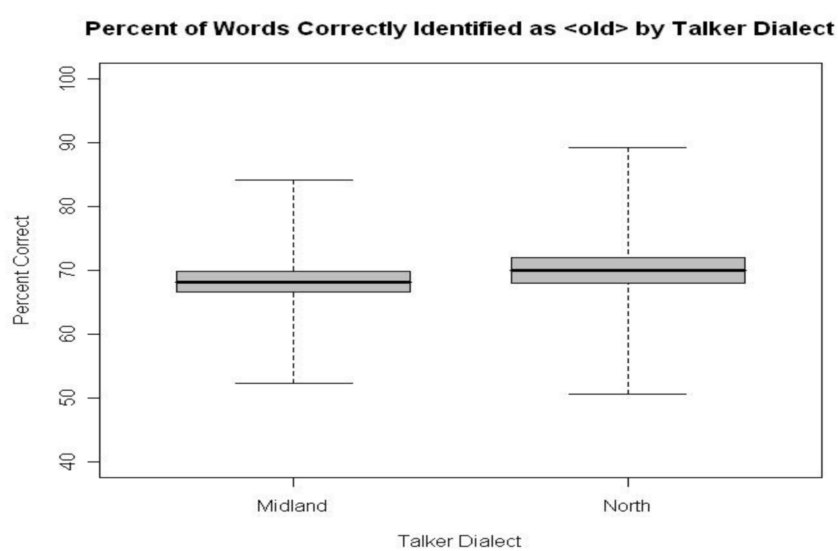

Figure 4: Percent of words correctly identified as "old" by the dialect of the talker in the test phase.

As shown in Figure 4, the listeners were slightly more accurate in recognizing "old" words when the talker in the test phase was from the Northern dialect region. However, the effect of dialect on word recognition accuracy in the test phase was not significant by a paired *t*-test.

Figure 5 shows the percent correct words recognized by the dialect of the original talker from the training phase (Midland or North) and repetition type (different-talker different-dialect, different-talker same-dialect, or same-talker).

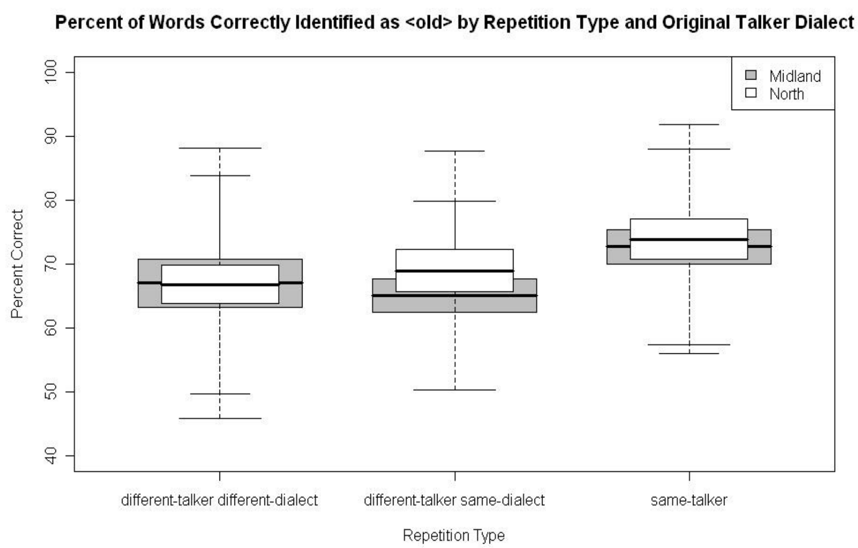

Figure 5: Percent of words correctly identified as "old" by the repetition type and status in the test phase.

As shown in Figure 5, listeners' were more accurate for same-talker repetitions than either different-talker different-dialect or different-talker same-dialect repetitions, which do not seem differ from one another. In addition, the listeners' performance seems to be the same for when the original talker from the training phase was from the Midland dialect region and when the original talker was from the Northern dialect region. A two-way repeated measures ANOVA on word recognition accuracy in the test phase with repetition type and original talker dialect as within-subjects factors was carried out. The main effects of repetition type and original talker dialect were not significant. The interaction of repetition type and original talker dialect was also not found to be significant. Paired *t*-tests revealed that accuracy for different-talker different-dialect and different-talker same-dialect repetitions did not significantly differ, but that accuracy for both different-talker different-dialect and different-talker same-dialect was significantly lower than accuracy for same-talker repetitions [$t(31)=-2.32, p<.05$ for different-talker different-dialect repetitions and $t(31)=-2.69, p<.01$ for different-talker same-dialect repetitions].

Word Recognition Response Times

Figure 4 displays the response times for hits in the test phase for each repetition type as a function of the dialect presented in the test phase. This figure shows that response times were generally slower for all repetitions when they were produced by a Northern dialect talker.

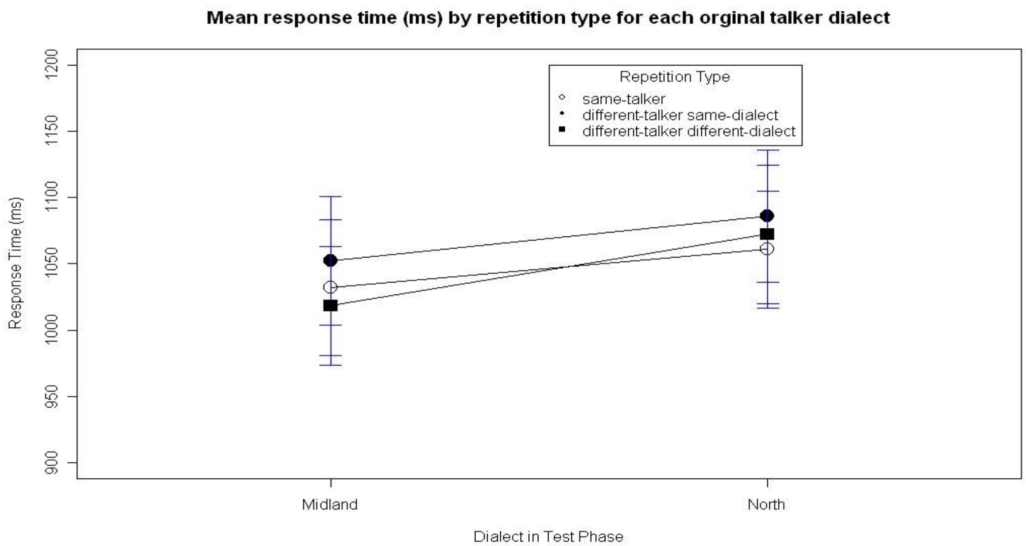

Figure 4: Response times for correctly recognizing old words for all repetition types by talker dialect in the repetition.

The main effect of the dialect on the repeated word was not significant by a t-test, but response times were also examined by original talker dialect when looking at the specific effects of the repetition type. Figure 5 displays the response times of hits for all repetition types by the dialect of the talker in the training phase (original talker dialect). As shown in the figure, when the original talker was from the Midland dialect region, response time was fastest for same-talker repetitions, slowest for different-talker different-dialect repetitions, and in between for different-talker same-dialect repetitions. However, when the original talker was from the Northern dialect region, response times were fastest for different-talker different-dialect repetitions, slowest for different-talker same-dialect repetitions, and in between for same-talker repetitions.

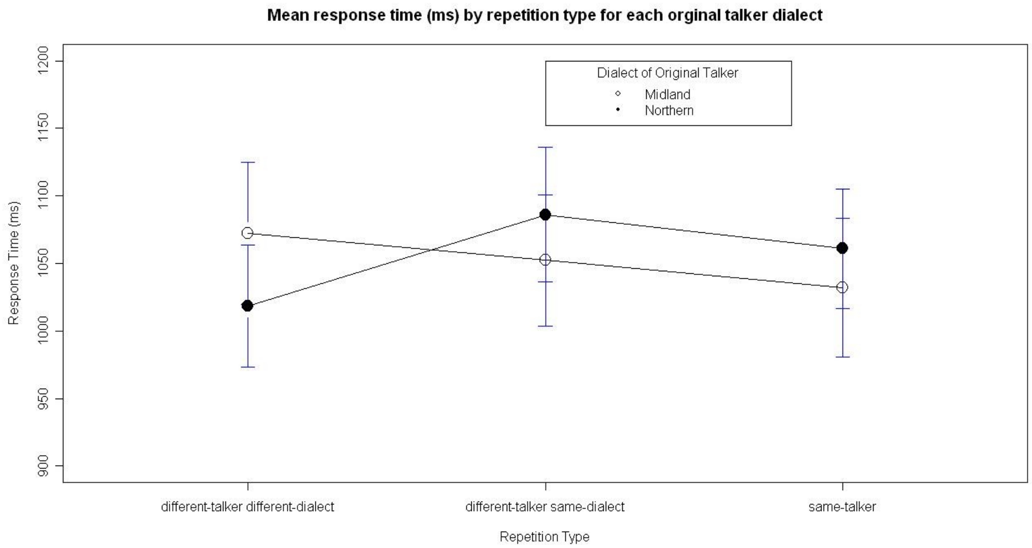

Figure 5: Response times for correctly recognizing old words for all repetition types by original talker dialect.

A 2 X 3 X 11 (Original Talker Dialect X Repetition Type X Vowel) ANOVA was carried out for the response times. The main effects of original talker dialect, repetition type, and vowel were not significant. The interactions of original talker dialect X repetition type and original talker dialect X vowel were also not found to be significant. The two-way repetition type X vowel interaction ($F(20,1382)= 2.8106$, $p<.001$) was significant. A significant three-way original talker dialect X repetition type X vowel interaction was found as well ($F(20,1382)= 1.9159$, $p<.01$).

The interaction of vowel and status was examined and three general patterns emerged in the response time data. Figures 6-8 display the three types of patterns found for vowels.

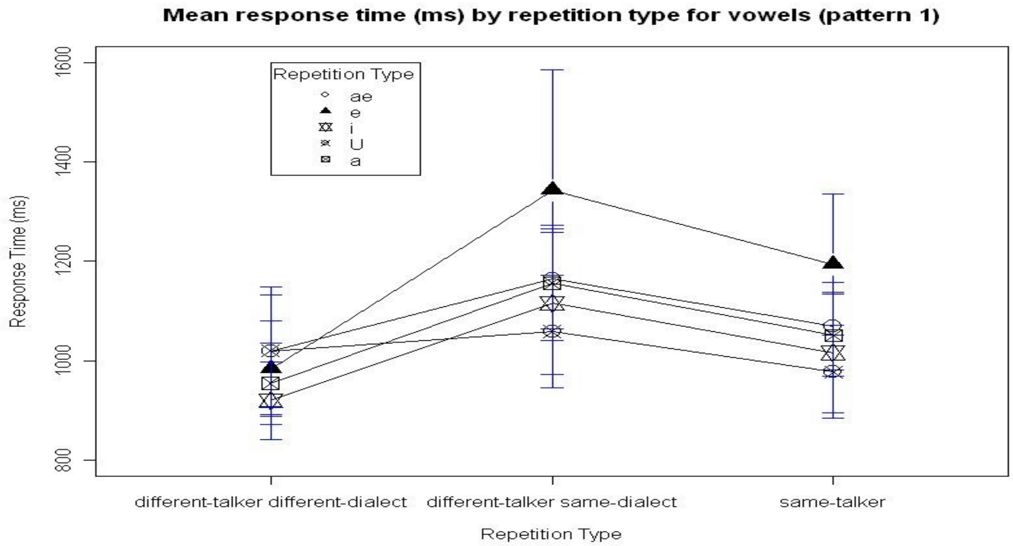

Figure 6: Response times for correctly recognizing old words for vowels that follow pattern 1.[1]

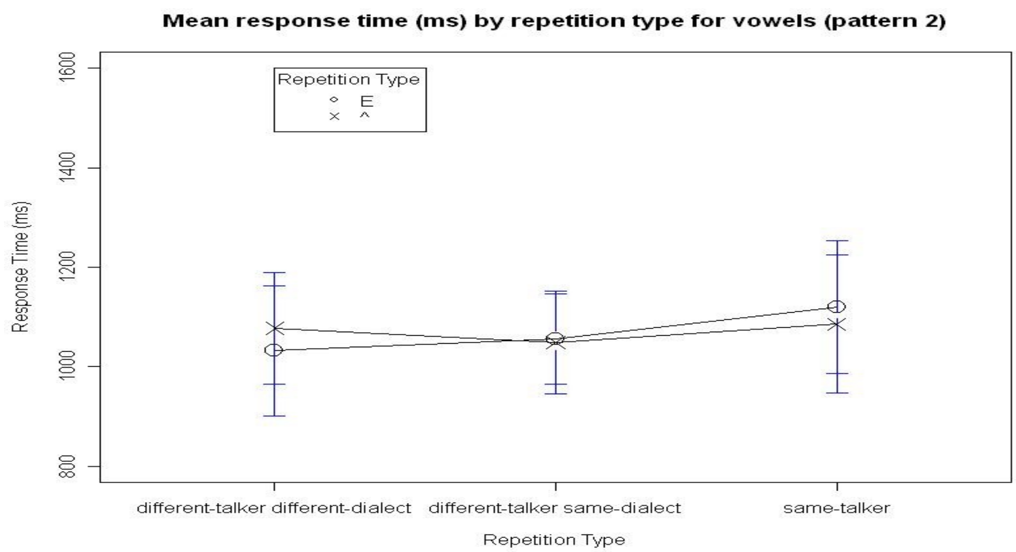

Figure 7: Response times for correctly recognizing old words for vowels that follow pattern 2.

---

[1]Note that in these figures the symbols ‘I’ represents /ɪ/, ‘E’ represents /ɛ/, ‘ae’ represents /æ/, ‘c’ represents /ɔ/, ‘^’ represents /ʌ/, ‘o’ represents /ow/, ‘i’ represents /i/, ‘u’ represents /u/, ‘ej’ represents /ej/, ‘a’ represents /a/, and ‘U’ represents /ʊ/.

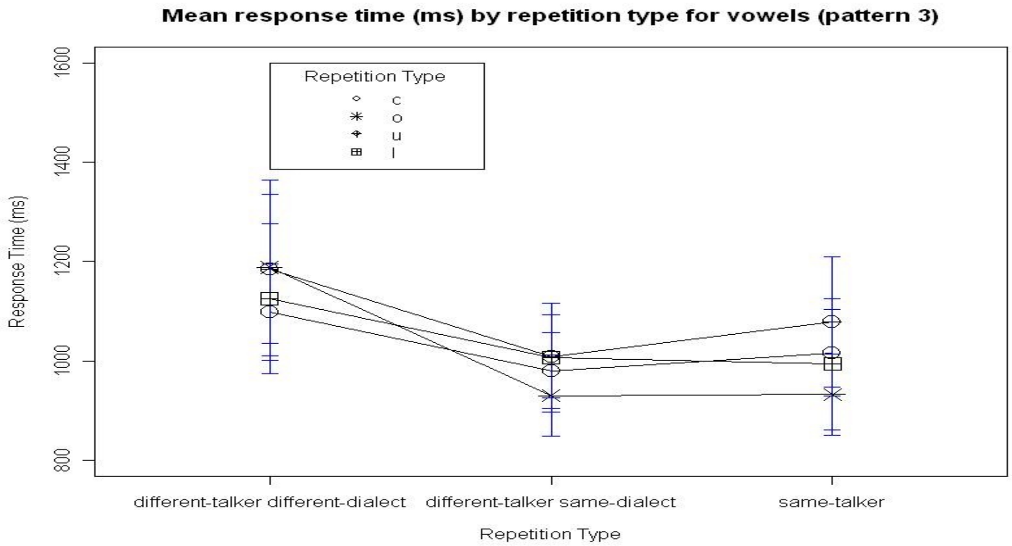

Figure 8: Response times for correctly recognizing old words for vowels that follow pattern 3.

As shown in Figures 6-8, vowels tended to follow three types of patterns. Vowels grouped in pattern 1 (Figure 6) show slowest response times for different-talker same-dialect repetitions and about the same response times for same-talker and different-talker different-dialect repetitions. Vowels grouped in pattern 2, as in Figure 7, show roughly the same response times for each of the three repetition types. Vowels grouped in pattern 3, as in Figure 8, show fastest response times for different-talker same-dialect repetitions, slightly slower times for same-talker repetitions, and slowest response times for different-talker different-dialect repetitions. To show the three-way interaction of original talker dialect X repetition type X vowel, Figures 9 and 10 give the response times for vowels for each repetition type for each original talker dialect.

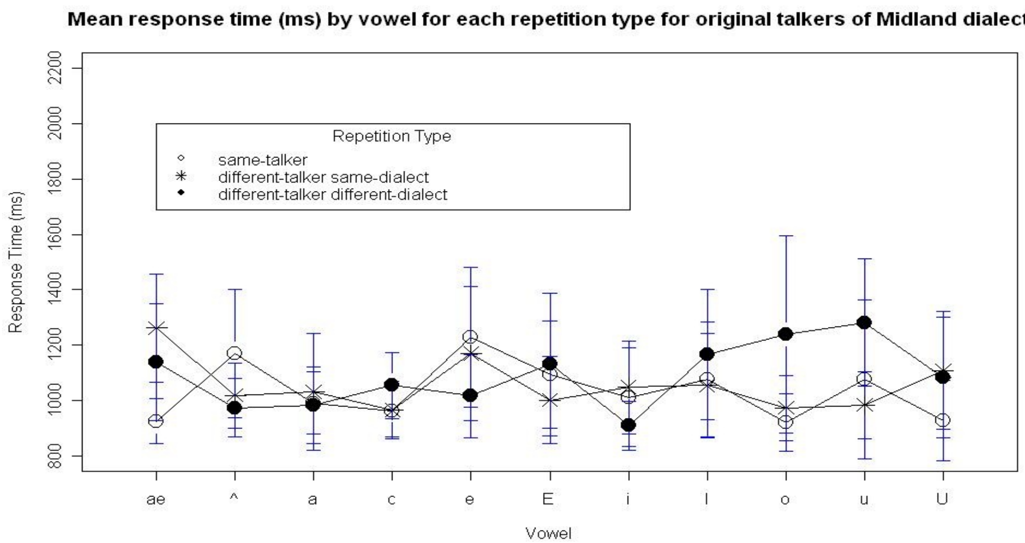

Figure 9: Response times for correctly recognizing old words by vowels for each repetition type when the original talker was from the Midland dialect region.

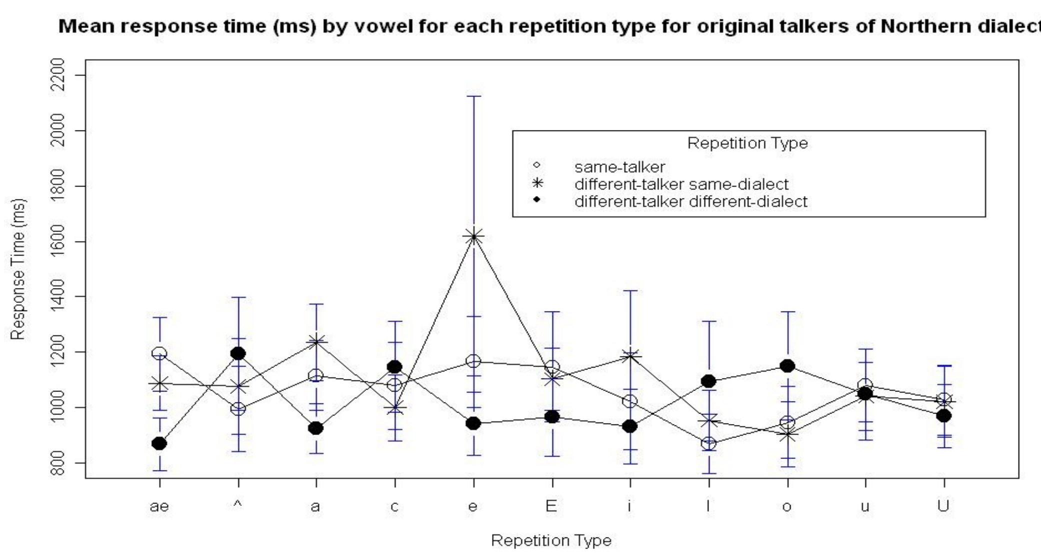

Figure 10: Response times for correctly recognizing old words by vowels for each repetition type when the original talker was from the Northern dialect region.

These figures show that the response times for vowels for the different repetition types are not the same for each original dialect. These interactions have not been further analyzed and, therefore, will not be discussed in detail.

Discussion and Conclusions

The results of the identification task from the training phase show that participants were significantly more accurate in identifying the target word when the talker was from the Northern dialect region than when the talker was from the Midland dialect region. However, performance for identifying the target vowel was not significantly different between talker dialects. These results are consistent with previous studies where participants were more accurate in identifying the target word for talkers from the Northern dialect region than the Midland dialect region in noise (Clopper & Pierrehumbert, 2007). Moreover, as in the current study, in Clopper & Pierrehumbert (2007), listeners did not show a signficant difference in accuracy between dialects for identifying the target vowel in noise. With this in mind, the results of the current study would be expected since the s and most of the same stimuli were used as in the Clopper & Pierrehumbert (2007).

The better performance for the Northern talkers, in general, is an interesting result. It would seem that participants should be better at identifying the target word and vowel from the Midland talkers since the Midland dialect is the least marked of the dialectal varieties of American English and shows very few dialect-specific features (Clopper, Pisoni, & de Jong, 2005). Because the results in this study were obtained for speech without noise, it likely rules out degradation due to noise, which was observed in the Clopper & Pierrehumbert (2007) study, as a cause for the better performance for Northern talkers. The better performance for Northern talkers for word responses and the same performance for both dialects for vowel responses, as in the Clopper & Pierrehumbert

(2007) study, suggests that participants misinterpreted consonants produced by the Midland talkers, responding with the incorrect word but correct vowel. To look at this in further detail, an acoustic study of these talkers is called for.

The results of the test phase show that participants were slightly more accurate for the same-talker repetitions than for both different-talker same-dialect and different-talker different-dialect repetitions. There was no significant difference between different-talker same-dialect and different-talker different-dialect repetitions. These results are similar to the Palmeri, Goldinger, & Pisoni (1993) findings for gender. In that study, there was no significant difference in word recognition accuracy between different-talker same-gender and different-talker different-gender repetitions. Still, because accuracy for the same-talker repetitions was significantly higher than both different-talker same-gender and different-talker different-gender repetitions, as in the Palmeri, Goldinger, & Pisoni (1993) study, this study seems to confirm the results of both Palmeri, Goldinger, & Pisoni (1993) and Goldinger (1996) that voice information much more detailed than gender, or, in this case, dialect, is stored in long-term memory and can facilitate word recognition. While this result might seem to suggest that participants are not coding for dialect, it may be that listeners rely on talker-specific voice information in word recognition tasks. That is, the word may be analyzed as either an exact voice match or not an exact voice match, in which case the extra information for dialect, or gender, would not result in a significant difference in accuracy.

Although the effect of the dialect of the talker on response time in the test phase was not significant, the data show that responses to stimuli produced by Northern talkers were slightly slower than responses to Midland talkers. This could possibly indicate that, although participants were more accurate in the training phase for Northern talkers, the Northern forms interfered with the recognition of the spoken words in the test phase. To look at the effect of dialect variation on the recognition task, then, the response times for each repetition type were examined separately by the dialect of the talker in the training phase (original talker dialect). As was initially predicted, when the original talker was from the Midland dialect region, recognition was fastest for same-talker repetitions, slowest for different-talker different-dialect repetitions, and in between for different-talker same-dialect repetitions. However, when the original talker was from the Northern dialect region, recognition was fastest for the different-talker different-dialect repetitions than either of the other repetitions, with responses to same-talker repetitions being slightly faster than responses to different-talker same-dialect repetitions.

The fast response time for the different-talker different-dialect talkers when the original talker was from the Northern dialect region might suggest that both abstract lexical representations and episodic traces are stored in long-term memory and contribute to perception. After hearing the original word in the Northern dialect, the processing of the same word in the Midland dialect is facilitated by a comparison to the stored episode and the abstract representation of the lexical item. With the vowel features of the Midland dialect being similar to the abstract categories, listeners are able to quickly retrieve the item. This interpretation would be consistent with hybrid exemplar-based models of speech perception, where both abstract lexical representations and episodic traces contribute to spoken word perception (Pierrehumbert, 2006). While the effect of repetition type and the interaction between repetition type and original talker dialect on response time were not significant, the resulting patterns in response time suggest that listeners code for dialect. This should be examined in further detail in future studies.

As mentioned above, the vowel X repetition type interaction and vowel X repetition type X original talker dialect interaction were all found to be significant. The interaction of repetition type and vowel tells us that response times were affected by repetition type differentially across vowels. As shown in Figures 6-8, the response times for the repetition types across dialects show three basic patterns. In the first pattern discussed, different-talker same-dialect repetitions are the slowest. In the second pattern, response times are the same for all repetition types. In the third pattern, response times are fastest for different-talker same-dialect repetitions and slowest for different-talker different-dialect repetitions.

The interaction of repetition type, vowel, and original talker dialect tells us that performance was affected differently across vowels for the different original talker dialect and repetition type combinations. Additional patterns arise when taking into account the original talker dialect, as in Figures 9 and 10. For example, for the vowel /ʌ/, same-talker repetitions are the fastest and different-talker different-dialect repetitions are the slowest when the original talker dialect was Northern but same-talker repetitions are the slowest and different-talker different-dialect repetitions are the fastest when the original talker dialect was Midland. This interaction would be expected since dialect-specific phonetic differences between the Northern and Midland dialects do not exist for all vowels (Clopper et al. 2005). On the same grounds, listeners would also be expected to react differently to each vowel based on its specific realization and the amount of similarity or dissimilarity that exists between that realization and the listeners' perceptual categories. However, the vowels that are exhibiting the patterns do not seem to fall into any natural groupings or categories that might be relevant to the Midland or Northern dialects. Still, the three-way interaction of repetition type, vowel, and original talker dialect does seem to suggest that dialect was affecting participants' performance, though differently for different vowels. Further analysis is needed to determine for which vowels this effect is most pronounced and if the effect is consistent with previous acoustic studies of the Midland and Northern dialects. The effect of the duration of the vowel should also be examined and taken into consideration for any future analysis of the vowel interactions.

The results of this study confirm that episodic memory traces of spoken words retain fine-grained surface details, as found in Goldinger (1996) and Palmeri, Goldinger, and Pisoni (1993). Though the results were not significant, this study found results similar to the results of those studies in that responses to same-talker repetitions were generally more accurate and faster than responses to different-talker same-dialect and different-talker different-dialect repetitions. In addition, the fast response times for the different-talker different-dialect talkers when the original talker was from the Northern dialect region suggest that both abstract lexical representations and episodic traces are stored in long-term memory and contribute to perception. Finally, the significant vowel interactions provide some evidence that dialect information is implicitly coded by the listener, though further studies are needed to better understand this result.

References

Clopper, C. G. & Bradlow, A. R.(in press). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech*.

Clopper, C. G. & Pierrehumbert J. B. (2007). Cross-dialect word recognition in noise. Presented at New Ways of Analyzing Variation (NWAV) 36, Philadelphia.

Clopper, C. G., Pisoni, D. B. & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustic Society of America, 118*, 1661-1676.

Clopper, C. G. & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32*, 111-114.

Clopper, C. G., Carter, A. K., Dillon, C. M. Dillon, Hernandez, L. R., Pisoni, D. B., Clarke, C. M., Harnsberger, J. D., & Herman, R. (2002). The Indiana Speech Project: An overview of the development of a multi-talker multi-dialect speech corpus. *Research on Spoken Language Progress Report No. 25* (Speech Research Laboratory, Indiana University, Bloomington), 367-380.

Fant, G. (1973). *Speech Sounds and Features*. Cambridge, MA: MIT Press.

Floccia, C., Girard, F., Goslin, J., & Konopczynki, G. (2006). Does regional accent perturb speech processing? *Journal of Experimental Psychology, 32*, 1276-1293.

Goldinger, S. D. (1996). Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166-1183.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics, 50*, 524-536.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustic Society of America, 97*, 3099-3111.

Labov, W. (1972). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Labov, W. & Ash, S. (1997). Understanding Birmingham. In C. Bernstein, T. Nunnally, & R. Sabino (Eds.), *Language Variety in the South Revisited* (pp. 508-573). Tuscaloosa, AL: Alabama University Press.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English*. New York: Mouton de Gruyter.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56*, 98-104.

Ladefoged, P. & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*, 98-104.

Palmeri, T. J., Goldinger, S. D., Pisoni, D. B. (1993). Episodic Encoding of Voice Attributes and Recognition Memory for Spoken Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309-328.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics. 34*, 516-530.

Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. In K. Johnson &

J.W. Mullinnex (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego: Academic Press.

Studebaker, G. A. (1985). A "Rationalized" Arcsine Transform. *Journal of Speech and Hearing Research, 28,* 455-462.

Summerfield, Q. & Haggard, M. P. (1973). Vocal tract normalization as demonstrated by reaction times. In *Report of speech research in progress* (Vol. 2, pp. 12-23). Belfast, Northern Ireland: The Queen's University of Belfast.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters: Part I. Recognition of backward voices. *Journal of Phonetics, 13*, 19-38.

Van Lancker, D., Kreiman, J., & Wickens, T. (1985). Familiar voice recognitions: Patterns and parameters: Part II. Recognition of rate-altered voices. *Journal of Phonetics, 13*, 39-52.

Winer, B. J. (1971). *Statistical principles in experimental design,* New York: McGraw-Hill.

Acknowledgments

This project was supported by the Colleges of the Arts and Sciences Undergraduate Research Award from the Ohio State University for 2007-2008 and the Undergraduate Research Award in the Humanities from the Ohio State University for 2007.